Crowdsourcing CPU vision Reading Assignment 1

Rupali Roy (UTEID - rur73)

1. What is the motivation for the authors in building the ImageNet dataset?

Ans. The authors believe that the enormous quantity of image data present on the internet today has the ability to support the development of more robust algorithm to index, organize retrieve and interact with the images that can be used for AI and deep learning. But the lack of proper organization of such data is a major problem. Thus the authors propose the idea of ImageNet dataset which will have millions of annotated images organized by the semantic hierarchy of WordNet.

2. What are the claimed contributions of the paper?

Ans. The paper proposes the idea of a new image database that is built upon the backbone of WordNet structure by populating the majority of the synset of WordNet with clean and full resolution images. The images present in the ImageNet have a higher level of accuracy and diversity. This accurate set of data is collected using the Amazon Mechanical Turk. The authors elucidate the use of ImageNet for visual recognition applications such as Non-Parametric Object Recognition, Tree-Based Image Classification, and Automatic Object Localization.

3. How were the images collected for the ImageNet dataset?

Ans. The images for the ImageNet dataset is collected in two steps, the first step involves collecting candidate images and the second step is cleaning the candidate image. The candidate images are collected by querying internet using different image search engines, to get maximum images we expand the query set by appending queries with a word from parent synsets and also translate the queries into other language. The second step is cleaning the candidate image by using Amazon Mechanical Turk.

4. How was the ImageNet taxonomy created?

Ans. ImageNet dataset is created on the hierarchical structure of WordNet. The WordNet is a lexical database which has approximately 80,000 noun synsets. The ImageNet database has millions of annotated images that is organized according to the semantic hierarchy of WordNet. It ensures that the images present in the database satisfy the required level of scaling, hierarchy, accuracy and diversity. Currently it has 12 subtrees with 5247 categories/synsets.

5. What approach was used to collect high quality image labels from crowd workers?

Ans. To collect high quality image labels from crowd workers Amazon Mechanical Turk is used. In AMT users label candidate images with the help of target synset definition. We do have some issues like

different opinions amongst the group of people or human error while labeling the candidate images, the best way to overcome these issues and get high quality image is to have multiple users label same images. The authors also mention an algorithm that dynamically determines the number of agreements needed for different categories of images, alongwith the confidence score table that gives an idea about the semantic difficulty of the synset. In order to get high quality labels for image we continue to the AMT user labeling until we reach a pre-determined confidence level.

6. How many categories are included in the final ImageNet taxonomy?

Ans. As mentioned in the research paper the ImageNet has 12 subtrees and 5247 categories. These categories have a total of 3.2 million images.

7. How many images are contained in the final ImageNet dataset?

Ans. As per the current statistics the ImageNet dataset has 12 subtrees with 5247 synsets containing 3.2 million cleanly annotated images in total. As mentioned in the paper, the authors believe that ImageNet will have a total of approximately 50 million images in coming two years.

8. Please submit one "discussion point". This can be in the form of a question, critique, or plausible future work that you think is interesting to investigate in greater detail in class. Possible topics for the discussion point include the proposed idea, methods, experimental design, and analysis of results.

Question- The paper discusses that the Automatic Object Localization uses the non-parametric graphical model for visual representation of objects against a global background class. I would like to discuss how the **non-parametric graphical model** is used to get the exact region of the object with maximum likelihood.